Artificial DNA: PNA & XNA 4:2, 37–38; April/May/June 2013; © 2013 Landes Bioscience

Digitizing humanity

Roy D. Sleator[1,*] and Aisling O' Driscoll[2]

[1]Department of Biological Sciences; Cork Institute of Technology; Cork, Ireland; [2]Department of Computing; Cork Institute of Technology; Cork, Ireland

The application of ex vivo synthetic DNA as a high capacity information storage medium is well documented. Herein, we consider the potential for synthetic DNA to be incorporated as part of the human genome; providing a definitive, accessible, in vivo database of patient history.

DNA, the original information storage molecule, comprising the biological script of life, may be the future of personalised data storage, enabling us to hard wire humanity.[1,2] A high capacity storage medium, with a theoretical storage potential of 455 exabytes per gram ssDNA,[3] DNA is not only extremely stable, it is also self-replicating, allowing information to be preserved and transmitted through the millennia.

Recent work by Church and colleagues[3] at Harvard described the conversion of html-coded data to DNA code using a 1 bit per base encoding (A,C = 0; T,G = 1); allowing the conversion of an entire book (*Regenesis: How Synthetic Biology Will Reinvent Nature and Ourselves* ISBN-13:978–0465021758) into DNA sequence. In an effort to reduce error and facilitate up-scaling, Goldman et al.,[4] described a modified strategy achieving a storage density of ~2.2 PB/g DNA. This modified approach first converts the original file type to binary code (0, 1) which is then converted to a ternary code (0, 1, 2) and in turn to the triplet DNA code. Replacing each trit with one of the three nucleotides different from the preceding one (i.e. A, T or C, if the preceding one is G) ensures that no homopolymers are generated – significantly reducing high throughput sequencing errors.[5]

While the above strategies focus on maintaining the DNA in vitro, in lyophilized form stored in a cool dark place, similar to archival magnetic tape, we suggest that in vivo storage may be a viable and in some cases more desirable option. Indeed, we envisage a scenario whereby all of the information relating to a particular individual; ancestry, health records, financial statements, criminal records etc. is encoded in the individual's own DNA. In this scenario, each time an individual visits a doctor a simple blood sample would contain all of the necessary information concerning the patient's health status, not merely based on the traditional biomarkers, but as rich readable files e.g., PDF's of family history, JPEG's of MRI scans etc. Furthermore, rather than the simple DNA fingerprint which we currently depend on,[6] blood from a crime scene would provide the image and address or the perpetrator, in addition to his or her previous indiscretions and full criminal history.

Disregarding for a moment the ethical, legal and social issues surrounding such an approach (which we acknowledge to be significant and complex),[7] let us consider the technical difficulties associated with such an endeavor: The first key difference between DNA stored in vitro and in vivo is that the latter is subject to replication and thus error accumulation during the copying process. One approach to reducing errors arising from spontaneous mutations is to introduce degeneracy; incorporating multiple copies of the desired sequence - multiple sequence alignment, following amplification, should generate a consensus sequence which is free of errors. Following this model, the more sensitive the data and the longer it is required, the more copies are necessary. Information which

Keywords: synthetic DNA, diagnostics, information, storage, identification

Submitted: 06/20/13

Accepted: 06/21/13

http://dx.doi.org/10.4161/adna.25489

*Correspondence to: Roy D. Sleator;
Email: roy.sleator@cit.ie

is only needed in the short-term e.g., specific infant dietary requirements i.e., first 12–18 mo of life, requires relatively few copies and can be allowed to mutate (become corrupted) over the first few years of life, with little negative impact on the relevant adult patient records.

A second major consideration is how to maintain the sequence information in vivo over the lifetime of the individual. One approach is to construct an entirely separate human artificial chromosome (HAC),[7] which can replicate autonomously in the nucleus of the host cells. While this approach facilitates a large storage capacity and physically separates the synthetic DNA from the host chromosomes, thereby reducing the potential for recombination events with the native DNA, a significant limitation is the absence of any selective pressure for the host cell to retain the synthetic chromosome. In the absence of such a selective pressure there is no physiological advantage for the cell to retain the artificial chromosome and the synthetic DNA will be diluted out of the system with each successive cell division. An alternative approach would be to tether the synthetic DNA to the existing host chromosomes which are naturally maintained by the host. Indeed, this approach could be extended such that medical information pertaining to genes residing on a particular chromosome is hardcoded to that chromosome (e.g., for cystic fibrosis sufferers, information relating to treatment history and prognosis would be encoded on Chromosome 7 on which the cystic fibrosis transmembrane conductance regulator gene resides).

The next major challenge facing our proposed approach of digitizing the human genome lies in our ability to transform the host cells—how do we get the DNA in there in the first place? One approach, germ line gene therapy, involves the transformation of germ cells i.e., sperm or eggs, with the synthetic DNA. This approach would in theory allow the transformed synthetic DNA to be heritable and passed on to later generations. However, a major disadvantage of this route is the lack of specific information pertaining to the individual themselves. Given that this route precedes the conception of the individual, the only information which can be incorporated would relate to the individual's ancestry. An alternative, though less stable approach, would involve transformation of pluripotent stem cells, specifically hematopoietic stem cells (HSCs) the progenitors of all the body's blood cells. Residing in the bone marrow, these cells can be extracted, expanded in vitro, transformed with synthetic DNA containing information pertaining to the individual and reintroduced back into the bone marrow. Following this procedure, all of the nucleus baring blood cells derived from the transformed HSCs will contain the synthetic DNA. Furthermore, owning to their self-renewal capacity, HSCs are thought to have indefinite lifespans (though this is likely determined by preprogrammed differences in repair capacity[8]) allowing the information to be maintained indefinitely.

Conclusions

While the aspirations of the current report are at best futuristic (and at worst fanciful), the underlying technologies are all well-established, making the proposed undertaking readily achievable. If realized, this approach elevates personalised medicine to a truly individual experience, revolutionizing traditional clinical diagnostics and treatment.

In a world of synthetic biology—resistance is futile…

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

RDS is and ESCMID Research Fellow and Coordinator of the EU FP7 IAPP Grant ClouDx-i. AOD is a ClouDx-i PI.

References

1. Sleator RD. Digital biology: a new era has begun. Bioengineered 2012; 3:311-2; PMID:23099453; http://dx.doi.org/10.4161/bioe.22367
2. O' Driscoll A, Sleator RD. Synthetic DNA: The next generation of big data storage. Bioengineered 2013; 4:123-5; PMID:23514938; http://dx.doi.org/10.4161/bioe.24296
3. Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA. Science 2012; 337:1628; PMID:22903519; http://dx.doi.org/10.1126/science.1226355
4. Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature 2013; 494:77-80; PMID:23354052; http://dx.doi.org/10.1038/nature11875
5. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE. Landscape of next-generation sequencing technologies. Anal Chem 2011; 83:4327-41; PMID:21612267; http://dx.doi.org/10.1021/ac2010857
6. Holobinko A. Forensic human identification in the United States and Canada: A review of the law, admissible techniques, and the legal implications of their application in forensic cases. Forensic Sci Int 2012; 217:222; PMID:22136971
7. Torgersen H. Synthetic biology in society: learning from past experience? Syst Synth Biol 2009; 3:9-17; PMID:19816795; http://dx.doi.org/10.1007/s11693-009-9030-y
8. Sieburg HB, Cattarossi G, Muller-Sieburg CE. Lifespan differences in hematopoietic stem cells are due to imperfect repair and unstable mean-reversion. PLoS Comput Biol 2013; 9:e1003006; PMID:23637582; http://dx.doi.org/10.1371/journal.pcbi.1003006